

3) LSTM Gates & Cell State

a) Roles of the Gates:

Gate	Activation Function	Purpose
Forget Gate	Sigmoid (σ)	Decides how much of the previous cell state to keep or forget (values close to 0 \rightarrow forget, close to 1 \rightarrow keep).
Input Gate	Sigmoid (σ) (for control) + Tanh (for candidate values)	Determines how much new information from the current input should be added to the cell state.
Output Gate	Sigmoid (σ)	Controls how much of the updated cell state is exposed to the hidden state (i.e., the output of the LSTM).

b) Linear Path for Gradients:

The LSTM cell state allows gradients to flow **additively (not multiplicatively)** across time steps, creating a near-linear path that helps preserve gradient magnitude and enables learning of **long-term dependencies**.

c) “What to Remember” vs. “What to Expose”:

- **What to remember** is managed by the **forget and input gates**, which decide which information to retain or update in the cell state.
- **What to expose** is controlled by the **output gate**, which determines how much of the internal memory (cell state) is revealed to the next layer or time step.